Do mnemonics help nurses learn melodic medical equipment alarms?

Alexandra N Wee
School of Information Technology and Electrical Engineering
The University of Queensland
St Lucia, Queensland, Australia

Penelope M Sanderson
ARC Key Centre for Human Factors,
The University of Queensland,
St Lucia, Queensland, Australia

We tested the melodic alarms and mnemonics proposed in the international standard for alarms used in medical electrical equipment (IEC 60601-1-8) for learnability and discriminability, with 14 registered nurses over two days, approximately a week apart. Two learning modes (Mnemonics and No Mnemonics) were examined in computer administered tests. The first experimental group was supplied with mnemonics while the second group was not supplied with a strategy. On Day 1, participants were introduced to the alarms. On Day 2, participants were tested and then allowed to relearn the alarms in a similar way to Day 1. The mean accuracy across all alarms other than the General alarm was poor (44% correct). Only one participant reached the learning criterion by the end of Day 2. Furthermore, participants with formal musical training identified the alarms with greater accuracy than those without (p=0.004). Our findings indicate that the identified problems need to be addressed before these alarms are introduced in practice.

INTRODUCTION

An important goal in developing medical equipment alarms is to ensure that alarms alert healthcare workers without startling them and their patients (Edworthy, 1994). In the medical domain the focus has moved away from designing loud and intrusive alarms, which might distress patients and distract medical staff, to designing audible and informative alerts to orient attention to areas where it might be important.

The latest International Electrotechnical Commission (IEC) standard for medical equipment alarms (IEC60601-1-8) includes many worthwhile recommendations for auditory alarms. The standard also recommends an optional melodic alarm set in which alarms are divided into physiological systems distinguishable by different melodies that may be less effective than desired. The seven physiological systems used are Cardiovascular (CV), Temperature (TE), Ventilation (VN), Perfusion (PE), Infusion (IN), Oxygen (OX) and Power Failure (PF). An additional General (GE) category exists for alarms which do not fit into any of the other categories such as failures to electrical or non oxygen gas supply systems. A medium and high priority version exists for each alarm.

To help distinguish between systems the alarms have been given distinct melodies and have been assigned a mnemonic to help operators learn and remember the nature of the device (Block, Rouse, Hakala, & Thompson, 2000). For example, the oxygen alarm is a descending series of notes, reflecting the decreasing levels of oxygen saturation seen a patient's blood if oxygen levels are inadequate, and the words to the alarm melody are "ox-y-gen al-arm".

Three previous studies by groups working independently from the standards committee have looked into the effectiveness of the IEC 60601-1-8 melodic alarms for alerting healthcare professionals to potentially dangerous events (Lacherez, Seah & Sanderson, 2006; Sanderson, Wee, & Lacherez, 2006; Williams & Beatty, 2005). The studies provide strong converging evidence that the melodies are not easy to learn and that some alarms are often confused with others.

Unfortunately, none of the three studies above is definitive and all could be improved to strengthen ecological validity. For example, Williams and Beatty (2005) used non-clinicians who had no experience with anaesthetic audible warning systems. Their instructions only included the words to the alarm melodies and did not provide the rationale for the melody as well. In the standard, only the rationale for the melody was included and not the words to the melodies which were presented only in Block et al. (2000). The IEC60601-1-8 standard would be the key document used to train medical practitioners, so the fact that Williams and Beatty did not provide the rationale is problematic. Williams and Beatty's results showed that participants had difficulty learning some alarms and confused many of them.

Sanderson et al. (2006) compared the performance of participants with and without access to the mnemonics and found very little benefit for mnemonics. The only observable difference between conditions was that the participants in the mnemonic condition were confused within several pairs of alarms whereas the participants in the non mnemonic conditions had a confusion pattern that was more idiosyncratic. However Sanderson et al. did not use trained medical volunteers in their study. Training was also limited to hearing each alarm only once per learning block. Lacherez et al. (2006) used registered nurses as participants but tested only the high priority alarms without the help of the mnemonics (many healthcare workers would learn the alarms without mnemonics) so the effect of the mnemonics with nurses is still unknown.

In the study described in this paper, we attempt to overcome some of the shortcomings of precious studies to see if performance will reach better levels. In particular, we test whether trained nurses are better able to connect the alarm mnemonics with their healthcare knowledge, and so show a benefit when they learn the IEC 60601-1-8 alarms with mnemonics rather than with no mnemonics–a benefit that was not seen for non-nurses in Sanderson et al. (2006).

Accordingly, in this experiment we compare the performance of registered nurses taught to identify the alarms using the IEC 60601-1-8 (2005) mnemonics with the performance of nurses who are not given mnemonics and are left to develop their own learning strategies. The nurses can listen to the alarms as many times as they feel necessary before starting a test session. In these ways, learning of the IEC 60601-1-8 alarms proceeds under the most favourable conditions of any study to date.

METHOD

Participants

Participants were 14 registered nurses from hospitals in the Brisbane Metropolitan area, recruited through flyers and word of mouth. Participants were paid for their participation with AUD$40 and a small gift. Although no formal hearing tests were administered, participants were asked prior to the experiment whether they had any known hearing problems.

Apparatus and Stimuli

Sixteen auditory warnings were created using Csound and conforming to the criteria outlined in the most recently published IEC standard for medical equipment (IEC 60601-1-8). High priority alarms consisted of five notes in a "N-N-N—N-N" pattern which lasted 2.25 seconds and repeated after a pause. Medium priority alarms consisted of three notes in a slower "N-N-N" pattern which lasted for 1.05 seconds. Sounds were processed on a Pentium® 4 1.9 GHz Acer Laptop with an integrated soundcard. Because we were interested in the learnability of the alarms at this point in the research program, and did not want the audibility of the alarms to be affected by background noise, the alarms were presented via AKG K 240DF Studio Monitor earphones at sound level that was comfortable and clear to the participant. Alarm labels and mnemonics were displayed on a 17 inch flat screen display at 1280 by 1024 screen resolution on a custom written Java program which also recorded the participants' responses.

Procedure

Participants were asked to complete a questionnaire which asked how long they had been in the nursing profession, their area of speciality and whether they played a musical instrument and for how many years. Participants were then introduced to the 16 alarms (8 medium priority, 8 high priority) over two days of testing. On day 1 the participants were randomly allocated into either the Mnemonic (M) group or the Non-Mnemonic (NM) group. Participants were seated in a quiet room and given a short introduction and instructions as to what was required of them during the experiment. The introduction consisted of a familiarisation period during which the participants were introduced to the alarm sounds and their associated labels. Participants in the M group were introduced to the mnemonics as memory aids for each alarm sound whereas the NM group are not given this information.

Day 1. Following the introduction, participants were allowed to learn the alarms. Buttons for all alarms were presented together on the screen. Participants were allowed to listen to the alarms as many times as they liked and in whichever order they liked. M condition participants were shown the Mnemonics in all learning screens. When the participant felt that they were ready to proceed they clicked on the button at the bottom of the screen to proceed to the testing screen. A randomised test set of alarms was presented to the participant one at a time with the participant attempting to identify the alarm by selecting from the 16 possible alarms.

After the test the computer showed participants a scorecard with their correct answers in blue and incorrect answers in red. When the participant had finished looking at their score they would go back to the learning screen to revise any alarms that they felt unsure about. This was followed by another testing session. This cycle of learning and testing continued until the participant achieved consecutive perfect scores or until the experiment time exceeded 45 minutes. No learning-testing stage was left incomplete.

Day 2. The second session for each participant took place between 6 and 9 days following the first session. When participants returned they were given a brief introduction to the day's experiment. The experiment was broken up into three stages, a long term memory test, a relearning phase and a transfer task. For the purposes of this paper we will only be discussing the results of the long term memory task and the relearning task.

Long term memory (LTM) testing. In the first stage the participants completed a series of tests much like the testing sessions they experienced in the first day. The only difference was that the test sessions were not broken up by learning phases. They were required to go through two test sessions with results presented at the end of each test session.

Relearning. Stage 2 of Day 2 was very similar to Day 1 of the experiment. Participants were once again allowed to learn the alarms via the learning screen which was followed by a test. This cycle was then repeated until the participant achieved two consecutive perfect scores or until after eight cycles were completed—whichever came first.

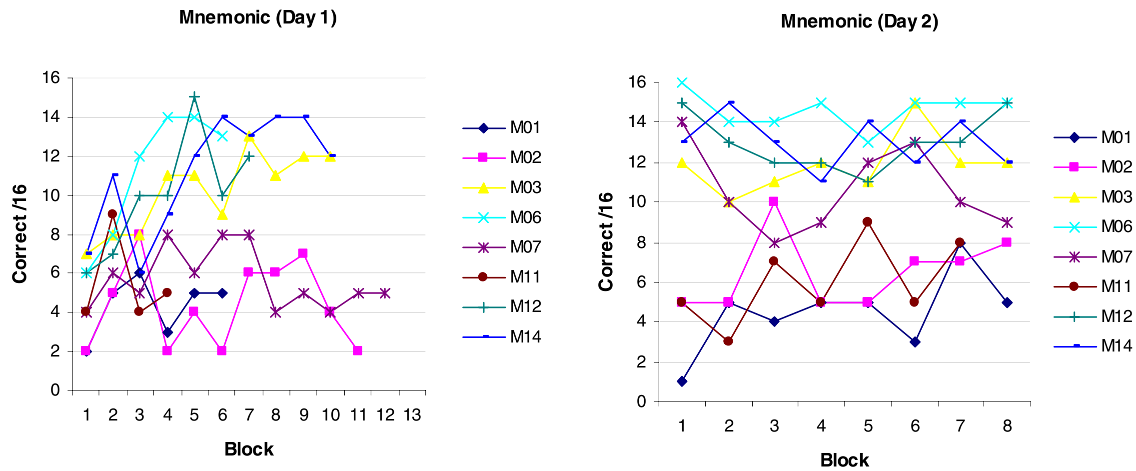


Figure 1 - Mnemonic condition. Learning curves for successive test results in the Mnemonic condition for each participant showing test block number and number correct in each test. Each line represents a different participant.

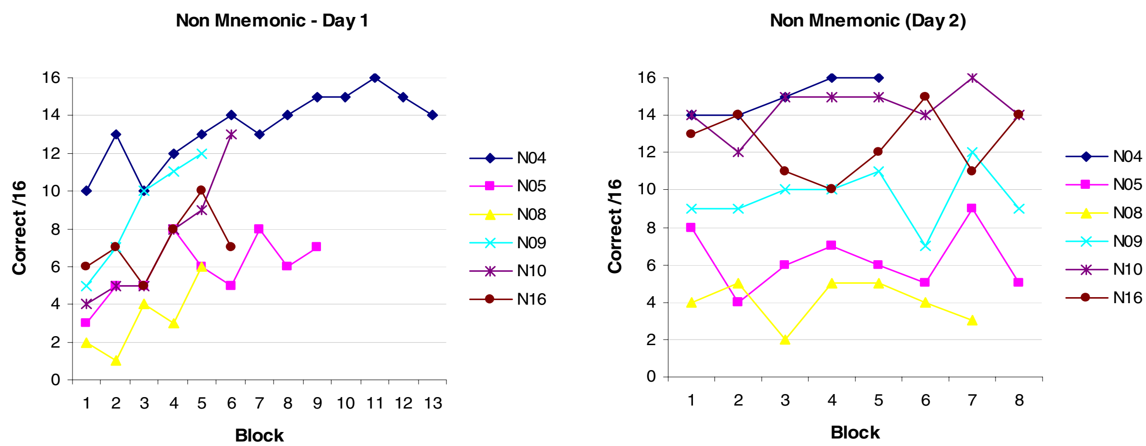


Figure 2 - Non-Mnemonic condition. Learning curves for successive test results for each participant in the Non Mnemonic condition showing test block number and number correct in each test. Each line represents a different participant.

RESULTS

Learnability

Day 1. The left graph in Figure 1 and Figure 2 shows performance on Day 1. Results are given for accuracy. With the exception of performance in identifying the GE alarm, which participants identified with a mean accuracy of about 93%, mean accuracy across the other alarms was poor at 44% (range 20%-61%) and is consistent with findings from all three previous studies. During the 45 minute experiment on Day 1, no participant achieved the learning criterion of consecutive perfect scores in the testing phases. Only one participant in the NM condition managed to complete one test with 100% accuracy but did not meet the learning criterion. Many participants showed a slow but progressing improvement whereas other participants showed little sign of improvement over the test blocks (Figure 1 and 2).

Day 1+2. Figure 1 shows results for the Mnemonic condition and Figure 2 for the Non-Mnemonic condition. At the end of the Day 2 trials only one participant managed to complete the learning criterion with two consecutive perfect test blocks.

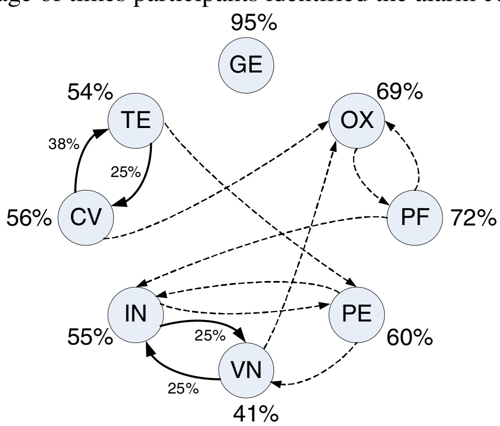
Looking at mean accuracy for the first two and last two blocks of Day 1, the LTM test at the start of Day 2, and the last two blocks of relearning on Day 2, there was a significant effect of learning phase, $F(3,30) = 4.12$, $p=0.015$, $MSE= 7.058$. A post-hoc Tukey HSD test showed that accuracy at the end of Day 1 and the end of Day 2 was greater than at the start of Day 1, $p = 0.032$ and 0.014 respectively. There was no observable effect of Mnemonics or No Mnemonics, $F(1,10) = 0.068$, $p=0.8$, $MSE = 18.21$.

There was a significant effect of Musicality on accuracy performance, with musically trained participants performing significantly better than those who reported no formal musical training longer than one year, $F(1,10)=13.635$, $p=0.004$, $MSE = 18.21$.

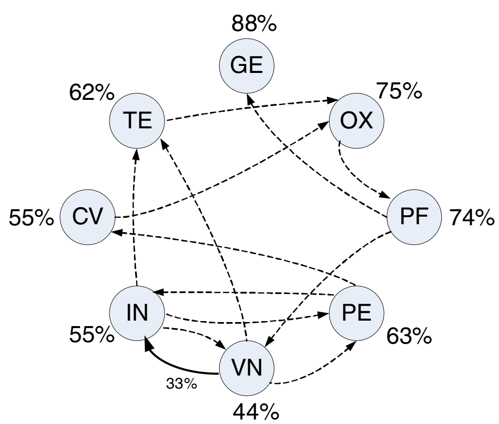
Confusions

Figure 3 and 4 shows the pattern of confusions for both the non mnemonic and mnemonic conditions at the end of

Day 2. The alarm at the origin of the arrow is confused for the alarm at the head of the arrow. The numbers on the links indicate the percentage of participants who on more than 25% of the test trials made the confusion. The dotted arrows indicate confusions that were unique to just one participant. The large percentages at each alarm node represent the percentage of times participants identified the alarm correctly.



(a) Confusions - Mnemonic (Day 2)



(b) Confusions - Non Mnemonic (Day 2)

Figure 3 – Confusions in (a)mnemonic and (b)non-mnemonic conditions. Numbers at each node indicate the percentage of correct answers for each alarm. Numbers on bold arrows indicate percentage of participants on more than 25% of the Day 2 tests confused the alarm at the start of the arrow with the alarm at the end of the arrow. Dotted lines represent confusions that were unique to just one participant. Clockwise: General(GE), Oxygen(OX), Power Failure(PF), Perfusion(PE), Ventilation(VN), Infusion(IN), Cardiovascular(CV), Temperature(TE)

DISCUSSION

This is the first study to test the IEC 60601-1-8 alarms in a way that compares the effect of the mnemonics supplied in the standard and that uses a sample of nurses who work in critical care areas and so are familiar with patient monitoring equipment. As in previous studies with non-healthcare workers (Sanderson, Wee & Lacherez, 2006) we found that the mnemonics provided in the IEC 60601-1-8 standard had no observable effect on the nurses' ability to correctly identify the alarm. Even so, only 1 of the 14 participants could correctly identify all the alarms to reach our test criterion after 2 days and that participant was in the non-mnemonic condition.

The confusion data indicate that certain groups of alarms are mutually confusable especially in the early stages of the experiment with some confusions persisting to the end of the experiment. From findings in Sanderson, Wee & Lacherez (2006) we expected to see more entrenched confusability between some of the alarms. Sanderson Wee & Lacherez (2006) speculated that the use of mnemonics increased confusion. We begin to observe some of these effects in the Mnemonic condition with 4 cases which more than one participant confused 2 alarms as opposed to only 1 case in the non mnemonic condition. However a larger sample size in this study would be necessary to test if this effect is reliable.

Previous findings that musical training is associated with alarm identification (Sanderson Wee & Lacherez, 2006; Lacherez et al., 2006) are replicated in this study, even with the relatively smaller sample size. Although we expect that musical training would be beneficial, the concern arises as to just how dependent the ability to identify alarms is on musical training. Thus designers of future alarms should aim to reduce the gap between the musically-trained and non musically trained population by designing alarms that are less dependent on special training.

All the studies conducted regarding the IEC 60601-1-8 melodic alarms were conducted after the publication of the standard. To our knowledge, no human factors studies regarding these alarms were conducted before initial publication of the standard in 2003.The results in this study as well as the three other relevant studies described in this paper strongly suggest that learnability and discriminability issues in the IEC 60601-1-8 melodic alarms should be addressed before being introduced into healthcare settings. Some of the problems identified in these studies would have been predictable and preventable using human factors evaluation techniques or simply based on previous literature on alarms in healthcare (Sanderson, Wee, Seah, Lacherez, 2006). Some key design issues such as designing alarms with different temporal patterns, keeping the number of different warnings used low (Edworthy, 1994; Edworthy & Hellier, 2005) and distinguishing on/off time between different alarms (Meredith & Edworthy, 1994) would have helped to develop a more recognizable and memorable recommendations for melodic alarms in the IEC 60601-1-8 standard.

REFERENCES

Block, F. E., Rouse, J. D., Hakala, M., & Thompson, C. L. (2000). A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. *Journal of Clinical Monitoring*, 16, 541-546.

Edworthy, J. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics*, 25(4), 202-210.

Edworthy, J., & Hellier, E. (2005). Fewer but better auditory alarms will improve patient safety. *Quality and Safety in Health Care*, 14, 212-215.

IEC60601-1-8. (2003-08-14). *Medical electrical equipment Part 1-8*. Geneva, Switzerland: International Electrotechnical Commission.

Lacherez, P., Seah, E., Sanderson P. (2006) Overlapping Melodic alarms are almost indiscriminable. *Human Factors.*

Meredith, C., & Edworthy, J. (1994). Sources of Confusion in Intensive therapy unit alarms. In N. Stanton (Ed.), *Human factors in alarm design* (pp. 207-219): Taylor & Francis, London.

Sanderson, P., Wee, A., Seah, E., & Lacherez, P. (2006). Auditory alarms, medical standards and urgency. *Proceedings of the International Conference on Auditory Displays (ICAD2006).* Queen Mary University of London, 20-23 June.

Sanderson, P., Wee, A., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia*, 61, 142-147.

Williams, S., & Beatty, P. C. W. (2005). Measuring the performance of audible alarms for anaesthesia. *Physiological Measurement, 26*, 571-581.